Defining lines of maximum probability for the design of patrol vessel itineraries

Fernando Bação[1]
Victor Lobo[2]
[1]ISEGI-UNL
Lisbon, Portugal
bacao@isegi.unl.pt
[2]Portuguese Naval Academy
Lisbon, Portugal
vlobo@isegi.unl.pt

SUMMARY

In this paper we deal with the problem of designing preemptive patrolling itineraries at sea. As a case study we use real data provided by the Portuguese Navy, concerning search and rescue missions. This can be seen as a curve fitting problem in which we aim at predicting future occurrences based on previous reported ones. We propose a method for the design of itineraries based on the use of a 1 dimensional self-organizing map. A number of tests are performed using self-organizing maps of different size. We conclude that the proposed method has a number of relevant characteristics such as robustness, speed and flexibility which are particularly adequate to the problem in hands.

KEYWORDS: *Traveling Salesman Problem, Route optimization, Prediction, Route planning*

INTRODUCTION

A large number of tasks performed by the Portuguese navy are related with non-military missions (NMM), such as enforcing the protection of the Portuguese Exclusive Economic zone from illegal fishing activities, monitoring environmental risks, and search and rescue missions (SAR). Here we will use de acronym NMM to designate the set of all of these missions. The definition of the best patrol route constitutes an important problem of the planning stage of these missions. Presently, this is done largely based on the previous empirical experience of the officer in charge and there are no guidelines or objective criteria to support the patrolling itinerary design.

It can be considered that the best route for a NMM is the route that roughly goes through all the points where occurrences of the same kind were reported in the recent past. Using this criterion the problem can be thought of as being an instance of the Traveling Salesman Problem (TSP), more rigorously a special case of the TSP. Although the analogy is correct it is important to emphasize that while in TSP the problem is essentially an operational one; here the focus is on producing the path that, based on the observed occurrences, "best predicts future occurrences". This way there is a major distinction between an operational problem like the TSP and the predictive nature of patrol vessel itinerary design (PVID). In PVID the major assumption underlying objective is that the best predictor for the future occurrences is given by the set of points that represent the position of the historically reported occurrences.

In this paper we seek to establish a strategy which enables a more consistent approach to the PVID problem in the context of NMM, establishing a more objective and optimal procedure. The fundamental idea is to propose a method to define the routes of "maximum probability" for preventing illegal fishing activities and allowing a closer proximity to areas of potential SAR missions. To a certain extent the task of defining itineraries can be seen as a special case of fitting a curve to a number of observations, much like what happens in regression analysis. The fundamental

difference resides on the fact that this is a non-parametric approach, where we use x,y geographic coordinates of past occurrences to obtain a line in that geographic space.The proceedings of AGILE2005 are the only written record of the conference, and are widely circulated. Therefore, it is important that the publication is of the highest possible quality. The quality and overall appearance of all submissions contribute to the general quality of the proceedings, and readers' perceptions of it. In order to avoid any reformatting later, also the abstracts should adhere to these guidelines.To summarise, we ask you to copy the format of this document exactly. You should use the same font, font size, line spacing and indents. This document exemplifies the formatting required.

RELATED WORK

Although there are important differences between the classic TSP and the PVID problem, the fundamental aspects of characterization, complexity analysis and possible solutions can be based on the TSP. The TSP is a well known problem that has been thoroughly studied in operations research (OR). The basic formulation (e.g. Baase and Gelder 2000) can be described as: if a salesperson has to visit n cities and if he desires to minimize the distance, which is the best path? Although the TSP problem is easily stated it has been a difficult challenge for researchers. The only way to find the optimal solution consists on calculating (at least implicitly) all the possible paths. The number of possible paths for n cities is given by n!. If the number of cities is small (5 or 10) it is perfectly feasible to search for all possible paths. Nevertheless, if the number of cities increases (for example 500) then it is simply impossible to search all possible solutions to the problem.

The TSP problem belongs to the class of NP-hard problems (Cook 1971; 1983). Proof of this can be found in (Papadimitriou and Steinglitz 1976). The fundamental consequence of this is that at present the best opportunity to solve the problem is by using heuristic methods. Although these methods do not guarantee optimal results they usually produce good solutions in adequate time. A good overview of the available heuristics to solve the TSP can be found in (Arora 1998).

The original problem formulation requires the algorithm to find a solution which passes rigorously through all the designated cities and the path to be closed. Clearly the problem we need to solve allows for the relaxation of this requirement, forcing the path to pass only near the cities, and allowing open paths. The relaxed problem was recently treated in (Dumitrescu and Mitchell 2003) using algorithms which required a specific neighborhood radius to be met. We propose the Self-Organizing Map (Kohonen 1982) as a simpler and more flexible approach to the problem. An explanation of the Self-Organizing Map is outside the scope of this paper, so we recommend Kohonen's latest book on SOM (Kohonen 2001) as the best reference on the subject.

The Self-Organizing Map (SOM) can be adapted to the TSP problem through the use of 1-dimensional maps. In this case the network is trained using the location of the cities that need to be visited. By the end of the training phase, the line representing the SOM will be laid out in such a way that it will pass close to areas with higher density of points. Using the SOM to solve the TSP problem is not new (Hueter 1988; Kim, Kim et al. 1993; Choy and Siu 1995; Sasamura, Ohta et al. 2002) nevertheless all previous approaches are focused on the original TSP formulation.

A PROPOSAL TO SOLVE THE PVID PROBLEM

One of the relevant aspects of this application is the fact that for the case of NMM the relative density of the points of previously reported occurrences is much more relevant than the actual exact position of each of the reported occurrences. Given the preemptive nature of these missions the objective is to use the historical data to loosely orientate the path, allowing for an optimal positioning of the vessel with regard to the areas where the probability of new occurrences is high. The fundamental idea is that areas of high density of reported occurrences constitute a good predictor of future occurrences. For this reason the areas of high density should be patrolled.

To accomplish these objectives we need a tool that is able to fit a line based on the density of previous occurrences. Taking advantage of some of the work developed in the adaptation of the SOM to the TSP we propose the use of a 1-dimensional SOM for the calculation of the PVID. In fact, the 1-dimensional SOM has some favorable characteristics which indicate that it can be a valuable option in this context. First, the 1-dimensional (1-D) SOM has proved to be quite capable when used in the context of the classic TSP. Secondly, it has been suggested that the SOM is robust in the presence of outliers (Cottrell, Fort et al. 1995; Openshaw and Openshaw 1997). Additionally, the magnification effect (Cottrell, Fort et al. 1998) has been quantified for the specific case of the 1-D SOM. This means that we are able to control the magnification effect of 1-D SOMs. Third, the SOM incorporates an important feature in the context of the PVID, which is the possibility of controlling the level of fitting of the itinerary based on the size of the SOM. Using SOMs with a smaller number of units the itinerary will progressively abandon the most distant points and concentrate on the areas of higher density of points. Using SOMs with a large number of units will lead to itineraries that closely match all past occurrences.

EXPERIMENTAL SETTING

All the data used in this study was kindly provided by the Portuguese Navy. Due to confidentiality and national security issues the data provided constitutes only a sample of the occurrences. The sample comprises 93 SAR occurrences during 2003. The initial sample included 7 additional SAR occurrences which were removed due to the fact that they were outside of what is known as the "Portuguese Strategic Triangle" (Mainland Portugal, Madeira Island, and the Azores) and were considered outliers. The database file was composed of x,y geographic coordinates and had the following format:

Longitude x_1	Latitude y_1
Longitude x_2	Latitude y_2
…	…
Longitude x_n	Latitude y_n

The final dataset is presented in Figure 1, were the circles represent SAR missions, and the lines represent the Portuguese coastline, Azores and Madeira.
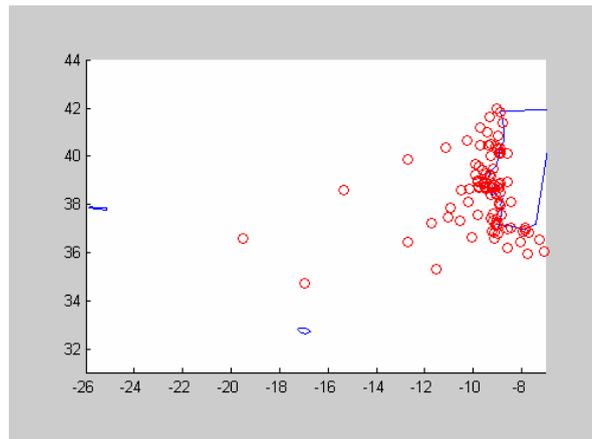

Figure 1: Locations of some SAR requests during 2003.

The software used to perform the tests was Matlab and "SOM Toolbox for Matlab" (Vesanto, Himberg et al. 2000). We test both open and closed paths as sometimes is not necessary to start and end the patrol in the same port.

The tests were performed with SOMs with different numbers of neurons. As the number of neurons decreases the path tends to disregard "outlier" occurrences concentrating on high density areas.

RESULTS

In the first set of tests we assumed that the patrol vessels should return to their home port, and so we used closed paths. As far as SOM goes, this requires that the units form a ring. We should force one of these units to be the home port, but in this case that is not necessary since there are very few possible home ports, and all of them are "visited" due to occurrences of SAR incidents in their vicinity. We ran experiments with a large number of SOM units (twice as many as data patterns), which lead to long and "winding" itineraries. We also ran experiments with fewer SOM units (93, 46, and 23 units), which lead to ever shorter and "straighter" itineraries. In each case, we ran 50 independent tests, which due to the non-deterministic nature of SOM lead to slightly different results. Examples of the itineraries obtained are presented in Figure 2, and the numerical results are presented in Table 1.
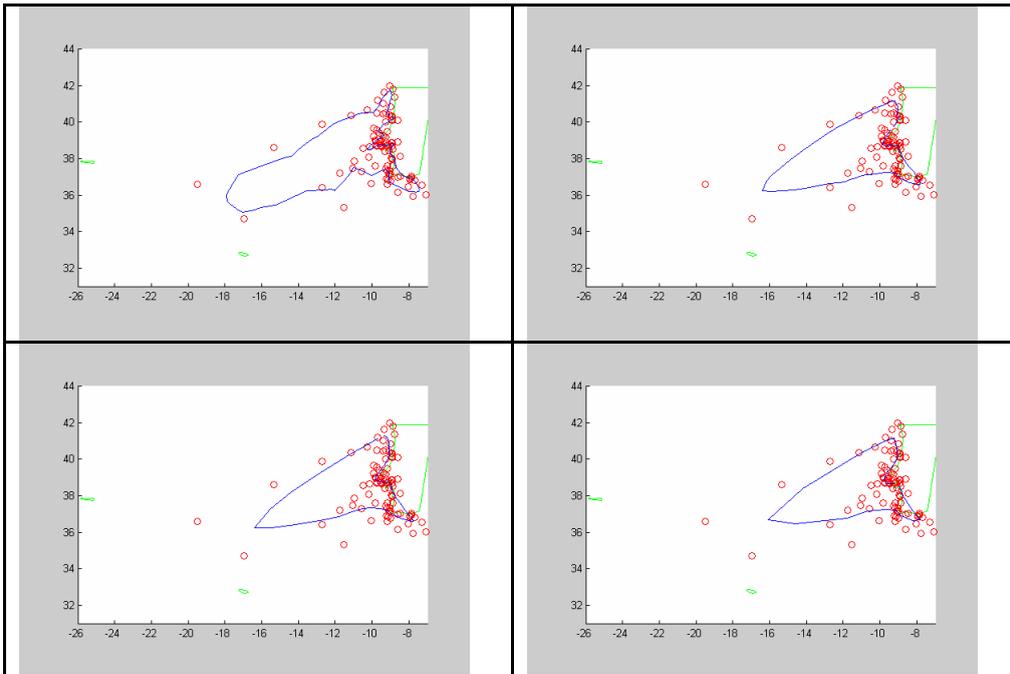

Figure 2: Trajectories obtained when we force closed itineraries (the vessel must return to it's home port). In the top left itinerary we use twice more SOM units (186) than available data points (186), in the top right we use 93 units, in the bottom left 46, and only 23 in the bottom right.

Number of SOM units	Average length	Standard Deviation
186	24.4130	2.5135
93	24.5072	2.6288
46	18.5052	1.2594
23	19.7034	2.3954

Table 1: Average distances (and standard deviations) obtained over 50 trials, using closed paths

We repeated all previous tests using open paths. In this case, it is assumed that the patrol vessels may start and end their cruises in different ports, which would be those closest to the beginning and end of the itineraries obtained. One again we ran 50 experiments for each number of possible SOM units, and obtained the results presented in Figure 3 and Table 2
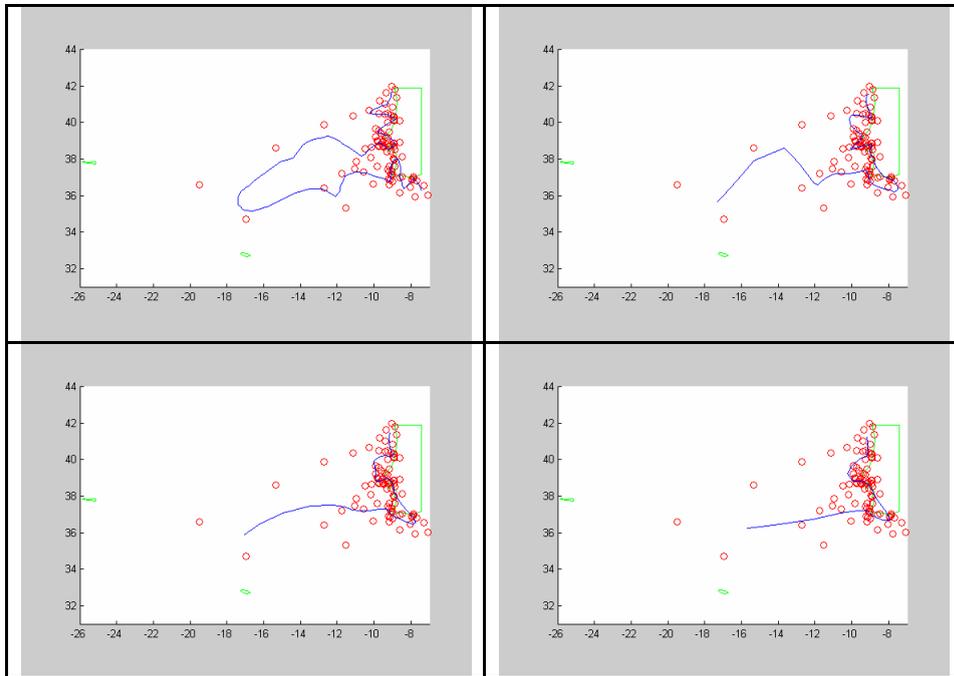

Figure 3: Trajectories obtained when we force open itineraries (the vessel start and end it's cruise anywhere). In the top left itinerary we use twice more SOM units (186) than available data points (186), in the top right we use 93 units, in the bottom left 46, and only 23 in the bottom right.

Number of SOM units	Average length	Standard Deviation
186	30.2446	0.2481
93	29.9153	0.3709
46	25.5010	0.3934
23	24.8238	0.5233

Table 2: Average distances (and standard deviations) obtained over 50 trials, using open paths

CONCLUSIONS AND FUTURE WORK

Based on our experiments we conclude that SOMs can be used as a flexible and quick way to solve the PVID problem. The processing time needed for the calculations is rather small (around 8 seconds on a standard PC) which allows real time redefinition of the path as new information is acquired. The use of points related with past occurrences gives the opportunity to design an itinerary which constitutes an informed "best guess" on where the presence of the patrol vessel is bound to discourage illegal fishing activities and be available for SAR missions. We do not, at present, show any comparisons with other methods, since no other systematic methods are used to perform this task.

Although the concept of using SOMs for this task was shown to be viable, some more work must be done to obtain good results. There are four main aspects that must be improved: the use of true geographic distances, forcing initial/final home ports, taking into account natural barriers, and using SOM unit density to define speed.

Throughout our tests, we used latitude and longitude (expressed in degrees) as Cartesian coordinates. This is in fact a crude approximation, especially since the latitudes involved are around 40ºN. In a first approximation, the true geographical coordinates could be projected onto a Cartesian plane that approximates the earth's surface in the area under study. The SOM could then use the standard Euclidean metric on these new coordinates, as was done in our tests. However, there would still be some distortion because of the necessarily non-linear projection. The best and only correct solution would be to use orthodromic distances throughout the calculations. While this poses no conceptual problem, it does require re-writing quite a lot of SOM code, and we suspect it would provide no substantial improvement.

If we whish to specify initial and final ports for the itineraries, we may force one (for closed paths) or two (for open paths) of the SOM units to have fixed coordinates. This is very easy to do, and we did run some tests using this approach. So as to make the results comparable we did not include them in this paper. This is not a very important issue, since in many cases it is better to suggest initial and final ports then to force them beforehand.

Some of the itineraries obtained in our tests cross land. This is obviously not possible for a patrol vessel (though it may still be useful for maritime patrol aircraft). One simple way of dealing with the problem is to sail between the two points where the itinerary crosses the shore by the shortest sea route. This approach, while simple, does imply that we are not using the best (shortest possible) itinerary. The correct solution to this problem would be to use true sea distances between points in the workings of the SOM algorithm. This may be achieved by calculating those distances using a Geographic Information System (GIS), which would also solve the problem of using Cartesian instead of geographic coordinates. Once again this would not affect the fundamental principals of our approach, but would require a lot of re-coding.

Finally, the density of SOM units is an indication of the density of past occurrences. Thus, areas where the density of SOM units is greater should be patrolled more carefully. The proximity of consecutive SOM units can be used to define the speed which should be used during the patrol. However, in many cases, it is better and more economical to use a constant speed. In this case, equally spaced target-times may be given to each unit, giving the vessel's captain the liberty to zig-zag or circle the locations of the units until those target-times are met.

BIBLIOGRAPHY

Arora, S., 1998, "Polynomial Time Approximation Schemes for Euclidean Traveling Salesman and Other Geometric Problems." Journal of the ACM 45(5), p 753–782.

Baase, S. and A. V. Gelder, 2000, Computer Algorithms, Addison-Wesley.

Choy, C. S.-T. and W.-C. Siu, 1995, New Approach for Solving the Travelling Salesman Problem using Self-Organizing Learning. IEEE International Conference on Neural Networks.

Cook, S. A.,1971, The complexity of theorem proving procedures. Annual ACM Symposium on Theory of Computing, p 151-158.

Cook, S. A.,1983, "An overview of computational compexity." Communications of the ACM 26(6), p 400-408.

Cottrell, M., J. C. Fort and G. Pages, 1995, "Comments about "Analysis of the convergence properties of topology preserving neural networks"." IEEE Transactions on Neural Networks 6(3), p 797-799.

Cottrell, M., J. C. Fort and G. Pages, 1998, "Theoretrical Aspects of the SOM algorithm." Neurocomputing, Elsevier 21, p 119-138.

Dumitrescu, A. and J. S. B. Mitchell, 2003, "Approximation algorithms for TSP with neighborhoods in the plane." Journal of Algorithms (Elsevier) 48(1), p 135-159.

Hueter, G. J., 1988, Solution of the traveling salesman problem with an adaptive ring. IEEE International Conference on Neural Networks, 1988, p 85-92.

Kim, S.-J., J.-H. Kim and H.-M. Choi, 1993, An efficient Algorithm for Traveling Salesman Problems based on Self-Organizing Feature Maps. IEEE International Conference on Fuzzy Systems, p 1085-1090.

Kohonen, T., 1982, Clustering, Taxonomy, and Topological Maps of Patterns. Proceedings of the 6th International Conference on Pattern Recognition, p 14-128.

Kohonen, T., 2001, Self-Organizing Maps. Berlin-Heidelberg, Springer, pp 501.

Openshaw, S. and C. Openshaw, 1997, Artificial intelligence in geography. Chichester, John Wiley & Sons.

Papadimitriou, C. H. and K. Steinglitz, 1976, Some Complexity Results for the Traveling Salesman Problems. Annual ACM Symposium on Theory of Computing - Proceedings of the eighth annual ACM, Hershey, Pennsylvania, United States, p 1-9.

Sasamura, H., R. Ohta and T. Saito, 2002, A simple learning agorithm for growing ring SOM and it's application to TSP. 9th International Conference on Neural Information Processing (ICONIP'02), p 1287-1290.

Vesanto, J., J. Himberg, E. Alhoniemi and J. Parhankangas, 2000, SOM Toolbox for Matlab 5. Espoo, Helsinki University of Technology, p 59.